

Media e deviazione standard

Secondo assegnamento Informatica CDL Fisica 2014/15

Data una serie di *misure* $x_1, i = 1 \dots N$, la media e la varianza campione (ossia le migliori stime della media e della varianza della distribuzione generatrice) si scrivono come

$$m = \frac{1}{N} \sum_{i=1}^N x_i \quad (1)$$

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - m)^2. \quad (2)$$

Esercizio 1. Il file `lotto_NA.txt` contiene lo storico dei primi numeri del lotto¹ usciti sulla ruota di Napoli a partire dal 1939 (sono esattamente 4218). Si scriva una funzione `media_dev` che legga i valori dal *file* (uno per linea separati da `\n`) e calcoli la media m e la deviazione standard s del campione utilizzando le equazioni (1) e (2) e stampando i valori risultanti sullo standard output. È un caso che la deviazione standard sia *vicina* al valore $89/\sqrt{12}$? E che sia ancora più vicina a $\sqrt{8099/12}$?

Esercizio 2. Sviluppando la (2) ci rendiamo conto che, attraverso alcuni semplici passaggi algebrici, essa può essere riscritta come

$$\begin{aligned} s^2 &= \frac{1}{N-1} \sum_{i=1}^N (x_i - m)^2 = \frac{1}{N-1} \left[\sum_{i=1}^N x_i^2 - 2m \sum_{i=1}^N x_i + \sum_{i=1}^N m^2 \right] = \\ &= \frac{1}{N-1} \left[\sum_{i=1}^N x_i^2 - 2Nm^2 + Nm^2 \right] = \frac{1}{N-1} \left[\sum_{i=1}^N x_i^2 - Nm^2 \right] = \\ &= \frac{1}{N-1} \left[\sum_{i=1}^N x_i^2 - \frac{1}{N} \left(\sum_{i=1}^N x_i \right)^2 \right]. \end{aligned} \quad (3)$$

¹Per chi non fosse familiare con il gioco si tratta essenzialmente dell'estrazione di numeri equiprobabili da 1 a 90 (compresi).

È un dato di fatto che il numero degli errori che si commettono nella stesura dei programmi ed il numero di bug è direttamente proporzionale al numero di operazioni ed al numero di linee di codice contenute nel programma stesso.

Usare la formulazione in (3) per implementare una funzione `media_dev2` con meno linee della funzione `media_dev`. Quali sono le principali differenze?

Esercizio 3. Sul pianeta Vulcano esiste un gioco del tutto simile al nostro—con la notevole differenza che i numeri estratti vanno da 1,000,000,001 a 1,000,000,090. Il file `lotto_VBA.txt` contiene uno storico delle estrazioni di una delle ruote di Vulcano (sono 4223). si confrontino i risultati ottenuti con `media_dev` e `media_dev2` su questo file e si commenti il risultato.

Esercizio 4. Si consideri il file `file_msr` che contiene una serie di misure (reali) una per linea separate da `\n` (sono esattamente 2356). Si scriva una funzione `media_dev_f` che calcola la media m e la deviazione standard s dei campioni contenuti in questo file.

Esercizio 5 Si scriva una funzione `max_min` che calcola il massimo ed il minimo dei numeri reali contenuti in `file_msr.txt`. Si scriva inoltre una funzione `media_dev_filter_f` che calcola la media m e la deviazione standard s del campione per i numeri contenuti in `file_msr.txt` scartando i k valori più piccoli ed i k valori più grandi.

Eseguire la funzione `media_dev_filter_f` per valori crescenti di k da 10 a 100 con passo 5. Che succede alla media ed alla varianza al crescere di k ?